TITLE: Mainstream Smoke Constituent Yields and Functional Relationships from a Worldwide Market Sample of Cigarette Brands. Part I: Smoke Constituent Yields at Standard ISO Smoking Conditions.

Authors: M. E. Counts*[a], F. S. Hsu[a], S. W. Laffoon[a], H. R. Wheeler[a], R. W. Dwyer[a], E. Kaelin[b], R. H. Cox[b]

a Philip Morris USA Inc., Research, Development & Engineering, Richmond, VA. 23261-6583

b Philip Morris International, Research & Development, c/o Philip Morris Products S.A., Quai Jeanrenaud 56, 2000 Neuchâtel, Switzerland

* Corresponding author.

ABSTRACT

The objective of this study was to evaluate a benchmarking approach for predicting smoke constituent machine yields for conventional cigarette brands from worldwide markets. Results for ISO smoke yields thus support the validity of a smoke constituent benchmark approach when brands, for which yields are to be predicted, have design characteristics within boundaries established by the exploratory brands. The use of conventional cigarette design parameters including tobacco blend, filter construction and ventilation, cigarette paper porosity, cigarette length and diameter, and processed tobacco components has led to a wide array of marketplace brand styles with different smoke yields. When cigarette design variants are coupled with regional agriculture and growing condition variations, the impact on smoke chemistry can be complex. Of the brands studied, the majority were "American blended" cigarettes containing blends of primarily bright, burley, and oriental tobaccos. In general, the yields of mainstream smoke constituents determined under standard ISO machine smoking conditions were well described by weighted least squares regression relationships with ISO tar ($R^2 > 0.80$ and coefficient p-values < 0.05). These relationships were validated with a subset of brands. Greater than 90% of the validation brands' smoke chemistry was within the 95% prediction intervals. Brands with constituent yields beyond the prediction intervals were primarily those with non-blended tobaccos. Additionally, regional agriculture and growing conditions may have contributed to smoke metals yield variation, as seen by the cadmium and mercury yields for two regional brands. The potential impact of

carbon in filters on vapor phase smoke constituent yields was addressed with multiple regression. Inclusion of a carbon factor variable improved the functional relationships between the yields of vapor phase constituents and ISO tar, as evidenced by regression statistics and reduced prediction errors. Functional relationships were developed for nitrogen oxides and tobacco specific nitrosamine smoke yields by including tobacco filler nitrate or tobacco specific nitrosamine concentration factors in the respective functional relationships. Average differences between measured and predicted yields were generally within the range of one to two measurement standard deviations.

Key Words: mainstream smoke constituents, smoke chemistry, benchmarking, prediction intervals, functional relationships, weighted least squares regression

Abbreviations: ISO, International Organization for Standardization; IEC, International Electrotechnical Commission; IARC, International Agency for Research on Cancer; CORESTA, Cooperation Center for Scientific Research Relative to Tobacco; ASTM, American Society for Testing and Materials; NNN, N'-nitrosonornicotine; NAT, N'-nitrosoanatabine; NAB, N'-nitrosoanabasine; NNK, 4-(N-methyl-N-nitrosamino)-1-(3-pyridyl)-1-butanone

INTRODUCTION

The tobacco industry has provided cigarette smoke yield tar, nicotine, and carbon monoxide data for various regulatory agencies for many years. More recently, regulations have been adopted or proposed for reporting the yields of specific smoke chemicals found in mainstream or sidestream cigarette smoke (Brazil, 2001; World Health Organization, 2000a; and International Digest of Health Legislation, 2001). Cigarette smoke is a complex chemical mixture. Nearly 4000 chemical constituents have been identified in mainstream cigarette smoke (Dube and Green, 1982, and Roberts, 1988). These individual chemical compounds and elements, as opposed to tar which is a composite of chemicals, are referred to as smoke constituents. Considerable research has been reported on cigarette smoke constituents identified as potentially toxic (IARC, 1986 and 1999; Reports of Surgeon General, 1989; Hoffmann, 1993, for example). Smith et al. (1997, 2000, and 2001) published comprehensive literature reviews of IARC Group 1, 2A, and 2B carcinogens found in mainstream cigarette smoke. Numerous publications are available that review standard smoking machine testing methods, alternative smoking methods, and the relevancy of any smoking machine testing protocol to smoking and health issues (World Health Organization, 2000b; Borgerding, 1997; Baker, 1999; National Cancer Institute, 1996 and 2001; Federal Register, 1997).

The reporting of individual smoke constituent yields in cigarette smoke raises a number of issues that remain to be resolved.. For example, there are no internationally recognized and standardized methods for the generation,

collection, and analysis of mainstream or sidestream smoke constituents. For smoke generation and collection, cigarette manufacturers in the United States are required to use the Federal Trade Commission (FTC) smoke method for tar, nicotine and carbon monoxide reporting (Federal Register, 1967 and 1980). Internationally, the International Organization for Standardization (ISO) method is recognized as the standard for determining cigarette tar, nicotine, and carbon monoxide yields (ISO 1999, 2000 a, b and c). Both the FTC and ISO standard methods are executed with smoking parameters of 35-mL puff volumes, 60-s puff intervals, and 2-s puff durations. Neither method incorporates other smoke constituent analysis guidelines. The Canadian Government has stipulated both machine smoking conditions and smoke constituent analytical methods in its Tobacco Act Reporting Regulation for cigarette products marketed in Canada (Canada, 2000a). This is the only known requirement for specific analytical methods within a regulation.

In contrast to tar, nicotine and carbon monoxide yield testing, considerable differences may exist among testing laboratories in practices for collection and analysis of smoke constituents (Purkis, 2001). This lack of standardized test guidelines can lead to greater variation in accuracy and precision than might otherwise be expected. A CORESTA Task Force Special Analytes Group was convened in 1999 with the objective of standardizing methodologies for the determination of smoke constituents through collaborative testing (CORESTA, 1999). This is the only known collaborative endeavor in which method standardization for smoke constituents is being internationally addressed.

Remove (D. Cox 5/19/03

Method standardization is a lengthy process requiring multiple laboratory participation. Comparing trace quantity smoke constituent yields from different laboratories remains a significant challenge without standard methods.

Laboratory availability and capacity are other issues to be recognized as the need for reporting of smoke constituent yields expands. Worldwide, there is a limited number of accredited laboratories with the capability and capacity to manage multiple smoke constituent analyses under a variety of smoking conditions for a large number of brands. Laboratory capacity may develop to meet demand if smoke constituent reporting regulations increase.

The objective of this study was to demonstrate the predictability of mainstream smoke constituent yields. An alternative to measuring all constituent yields for each brand in a market is to develop functional relationships between the yields of standard cigarette yield parameters and smoke constituents. In this study, relationships were developed from one representative sample set of cigarette brands and then applied to a test series of brands. Benchmarking, as used here, encompasses the testing of a set of cigarette brand samples having cigarette design features that are representative of the marketplace. Functional relationships are then established between the yields of each smoke constituent and the predicting variables. The predicting variables are generally the fundamental machine yield parameters of mainstream smoke tar, nicotine or carbon monoxide determined using standardized smoking methods. These functional relationships were subsequently used to predict constituent yields for other brands that were not tested.

This study was not the first to demonstrate the potential utility of a benchmarking approach. It was unique, however, in that the study included brands marketed in a wide distribution of geographical regions. Benchmarking offers a technical alternative in providing reliable smoke constituent yield information. This article reports benchmarking results at standard ISO smoking machine conditions.

There are numerous published examples of the use of fundamental yield parameters to predict, compare, or rank other smoke chemistry yields. Several studies included machine smoking conditions other than FTC or ISO. Young et al. (1981) evaluated the adequacy of tar for predicting carbon monoxide, hydrogen cyanide and acrolein. Their conclusion at that time was that tar was an inadequate predictor variable for these constituents. In a later study, Rickert et al. (1983) compared tar, nicotine, and carbon monoxide yields of 26 brands, with and without filter ventilation blocking, at several smoking machine conditions. They concluded that the relative ranking of cigarette yields at three smoking conditions, with 100% filter ventilation blocking at the "intense" condition, was essentially maintained. Tar, nicotine, and carbon monoxide at two alternative smoking conditions could also be predicted from yields at standard smoking conditions. Jenkins et al. (1983) investigated mainstream smoke yields of tar, nicotine, carbon monoxide, nitrogen oxides, and hydrogen cyanide at FTC smoking conditions for 32 cigarette brands over a wide tar range. They found that the smoke yield of the constituents studied could be estimated within 50% if the brand's tar yield were known. Chepiga et al. (2000) reported mainstream

smoke constituent yields for a 1995 sampling of 29 U.S. market cigarette brands and two University of Kentucky reference cigarettes. The analytical work was part of a study to assess smoke chemistry and smoke condensate mutagenicity relationships for commercial brands and reference cigarettes. For all brands tested at FTC smoking conditions, mainstream smoke constituent yields increased with increasing tar. Linear correlations between mainstream tar and constituent yields were significant. The authors noted that this was particularly relevant because of the wide range of design features of the cigarette brands studied. Seeman et al. (2003) compared mainstream tar and acetaldehyde for over 100 U.S. commercial brands in each of several years between 1985 and 1993. FTC mainstream smoke acetaldehyde was significantly correlated with both tar and carbon monoxide in each year's sample.

An extensive benchmark study was done cooperatively with the Massachusetts Department of Public Health (MDPH) by four U.S. tobacco companies in 1999 (Borgerding et al., 2000). Mainstream smoke constituent yields of brands representative of the U.S. market were determined at smoking conditions defined by the MDPH: puff volume, interval, and duration of 45 mL, 30 s, and 2 s, respectively, with filter ventilation 50% blocked (Massachusetts, 1997). Functional relationships were developed between tar, nicotine, and carbon monoxide, as measured at either standard FTC or MDPH-defined smoking conditions, and individual smoke constituent yields determined at MDPH-defined conditions. These relationships had the potential for predicting smoke constituent yields for other similarly designed brands.

An indirect example of benchmarking is contained in the Canada Tobacco Act Reporting Regulation. The regulation allows for use of linear functional relationships between tar or nicotine and each of the smoke constituents specified in the regulation (Canada, 2000b). Prediction models could then be used for smoke chemistry yields for either ISO or Health Canada alternative smoking conditions specified in the regulation.

An important element of the benchmarking approach is that brands for which smoke constituent yields are to be predicted must have design parameters that are within the range of the design parameters of the brands selected to represent the market. Design parameter ranges include various tobacco blends, weights, filter ventilations, and cigarette paper porosities, for example. The study reported here and the particular functional relationships developed are known to be applicable solely to current and future brands of Philip Morris USA Inc. and Philip Morris International Inc. that are within the design range defined by the study cigarettes. The study approach and the statistical treatment of data are applicable to other studies of cigarette mainstream smoke constituent yields.

MATERIALS AND METHODS

Cigarettes

Philip Morris commercial brands from various international market regions were tested. Forty-eight filtered brands were selected to include the majority of cigarette design features and tar categories available in those regions. Most brands were blended cigarettes consisting primarily of bright ("Virginia" or flue-cured), burley (air-cured), and oriental tobaccos. Three brands contained

primarily bright tobaccos. Nine brands included 14 to 50 mg of activated carbon (also referred to as activated charcoal) in their filter construction. ISO tar yields ranged from approximately 1 to 14 mg per cigarette. Brands were sampled at 14 production facilities in late 1999. Table 1 is a summary of characteristics for these cigarette brands. The study used a split-sample analysis plan. Forty cigarette samples, including 39 commercial brands and the University of Kentucky 1R4F reference cigarette (Davis et al., 1984) made up the exploratory sample group. The other 9 commercial brands and the 1R4F reference cigarette were in the validation sample group. The validation group included a range of tar yields and region representation, and was tested after completion of testing of the exploratory group. The exploratory group data were used to develop functional relationships. The validation group data provided independent assessments of the predicting strength of the functional relationships. The 1R4F reference cigarette was included with each sample group for consistency comparisons.

Mainstream Smoke Collection Methods

The study was performed at ISO smoking machine conditions of 35-mL puff volume, 2-s puff duration, 60-s puff interval. Cigarette samples were conditioned following ISO Standard 3402 (ISO 1999). Twenty replicates were analyzed for tar, nicotine, and carbon monoxide yields. Seven replicates were analyzed for mainstream smoke constituent yields.

Mainstream Smoke Constituent and Tobacco Filler Analyses

An ISO/IEC 17025-accredited commercial laboratory conducted the machine smoking and constituent analyses. The constituents (Table 2) analyzed for this study were representative of those included in several current or proposed government regulations (Brazil, 2001; Canada, 2000a; and Massachusetts, 1998). Analytical methods used by the laboratory were the same as those found in current Health Canada regulations (Canada, 2000b). Tobacco filler from each cigarette brand was also analyzed for nitrate and tobacco specific nitrosamines (TSNA) NNN, NAT, NAB, and NNK. Outlier statistics (ASTM, 1994) were applied to smoke constituent yield and filler chemistry data. Smoke constituents analyzed, the number of cigarettes per replicate, and the respective Health Canada mainstream smoke and whole tobacco analysis methods are found in Table 2.

Statistics

A functional relationship is a mathematical equation that specifies the quantitative relationship between a dependent variable and one or more independent variables. A statistically significant functional relationship does not prove cause and effect, but only establishes the degree of association between the dependent and the independent variable(s). To prove causality, the investigator must conduct additional studies specifically designed for the purpose (Kleinbaum et al 1988). This equation may then be used for predicting values of a dependent variable when values of independent variables are known. Regression is the computational process used to develop functional relationships

from observed values of the dependent and independent variables. A key factor in developing functional relationships for smoke constituents was the selection of representative cigarette brands for the exploratory sample set. These brands and their design features established the boundary conditions for future prediction of yields for other brands. Valid predictions could be made only for brands with design characteristics within the boundary of the exploratory set. The ranges of these design features that determine the boundary conditions are summarized in Table 1. Extrapolation beyond these ranges is misuse of the functional relationships. A second factor in the benchmark approach was selection of the independent variables from which strong functional relationships can be developed. The independent variable or variables should be readily determined through validated methods. Mainstream tar, nicotine, and carbon monoxide yields at ISO conditions were suitable independent variable candidates for functional relationships.

Relationships between mainstream smoke tar, nicotine, and carbon monoxide and individual smoke constituents were investigated using SAS statistical software (Release 8.01, SAS Institute, Cary, NC.) and EXCEL 97 (Microsoft, Redmond, WA.). The coefficients of determination, R^2 and adjusted R^2, together with other diagnostic parameters were used to assess the strength and statistical significance of each functional relationship. These diagnostics included p-value and F-statistics for regression coefficients, root mean square errors (RMSE), and predicted residual error sum of squares (PRESS) (Myers, 1990, and Glantz et al., 1990). Each statistic quantifies different aspects of the regression relationship

between independent and dependent variables. The larger the value of R^2 and the lower the values of RMSE and PRESS, the better a particular relationship describes the association between independent and dependent variables and, therefore, the stronger its potential predictive utility. Simple linear, quadratic, and multiple regression approaches were compared. Regression with transformed variables was also evaluated for several smoke constituents that were weakly correlated with tar, nicotine, or carbon monoxide alone. These transformed variables incorporated a factor for filler chemistry. As will be seen, this approach was used primarily for the yields of mainstream smoke nitrogen oxides and tobacco specific nitrosamines.

Functional relationships were assessed for their capacity to predict the smoke constituent yields of the validation brand set, as manifested by prediction errors (PE). These are the differences between predicted and measured smoke constituent yields. The more representative a functional relationship is of the measured values, the lower the prediction errors. It is recognized that there are no particular criteria for determining the acceptability of a functional relationship. It is a judgment based on the evaluation of the statistical parameters and the regression plots.

RESULTS

Mainstream ISO Smoke Constituent Yields

Average mainstream smoke tar, nicotine, carbon monoxide and smoke constituent yields for each brand in this study and the 1R4F reference cigarette are found in Table 3. The results are grouped by exploratory brand and validation brand sample sets. Several smoke constituent yields, particularly for lower tar yield brands, were below the limit of detection (LOD) or limit of quantification (LOQ) established by the contract laboratory. Those constituents and respective LOD's and LOQ's are listed in Table 4.

Functional Relationships

(1) Regression Approach Evaluation

Functional relationships for each mainstream smoke constituent were generated by regressing the mean yields of the 39 exploratory brands with the mean yields of tar, nicotine, or carbon monoxide, as measured at ISO smoking conditions. For the purpose of functional relationship development, analytical limits for LOQ and LOD were used instead of omitting low constituent yield brands from the analyses. Brands with higher yield values would therefore not impart larger influences because brands with yields lower than measurement limits were omitted from the regression analyses. Regression coefficients, diagnostic statistics and residuals analyses were examined. There were no substantive differences in linear and quadratic regression statistics. Therefore, linear models were selected.

E. Kaelin 5/21/03

See note #1

Replace with: In the quadratic model, x^2 coefficients were always not statistically different from 0.14 Therefore, linear models were selected.

One of the basic assumptions for regression is that residuals are randomly distributed over the range of the independent variable (i.e., homoscedasticity) (Kohler, 1994). However, for most of the smoke constituents, residuals tended to increase with increasing tar, nicotine, or carbon monoxide yield. This indicated that many constituents failed to meet the regression criterion for homogeneous and normally distributed variance.

Several approaches can be used to stabilize the variances. These include, for example, data transformation or weighted least squares (WLS) regression. The approach eventually used is frequently determined by trial. WLS was selected for this study. WLS is a modification of standard regression analysis. Once the proper weight factor (w_i) is determined that can best correct for the unequal distribution of variances, the methodology then determines the regression coefficients (a, slope and b, intercept) by minimizing the sum of the squared residual, SS_{res},

$$SS_{res} = \sum_{i=1}^{n} w_i (y_i - b - a x_i)^2$$

where w_i = weighting factor, y_i = dependent variable and x_i = independent variable (Myers, 1990, Ch 7). Data with large error variances will contribute less to the SS_{res} with weighting. Weighting is different from normalization in which the constituent yields are divided by a user-defined constant. In WLS, weighting is not applied directly to the constituent yields but only to the residuals, as shown in the above equation.

Different weighting factors were evaluated. The inverse of constituent variance, (i.e., $1/\sigma^2$), as frequently used, resulted in prediction intervals that decreased, but also varied irregularly in width, over the range of tar or carbon monoxide studied. Those brands with constituent yields below LOD or LOQ could not be weighted because of the absence of standard deviations and therefore were eliminated by this $1/\sigma^2$ WLS approach. An alternative approach was using the inverse of the independent variable, e.g., 1/tar or 1/CO, as the weighting factor which was generally proportional to the constituent variances. This resulted in randomly distributed residuals, increased R^2, and reduced RMSE and PRESS statistics as compared to the unweighted analysis.

The resulting weighted regression slope and intercept, generally less sensitive to the weighting, were very similar to those from the unweighted analysis. However, the improved regression statistics were reflected primarily in the error variances, confidence intervals, and prediction intervals (Miller et al., 1993 and Mandel, 1991). WLS generated prediction intervals with nonconstant width that expanded at higher yields and contracted toward the low yield region. The varying width was more proportional to the constituent yield dispersion over the tar range. The prediction intervals were therefore believed to be more appropriate for the data used to develop the functional relationships in this study. Weighted regression with inverse of independent variable weighting was adopted for the development of functional relationships.

The regression parameters from WLS are shown in Table 5. Regressions were not included for nickel, chromium, selenium, and arsenic since their ISO

smoke yields were below the limits of detection or quantification for the majority of brands.

(2) Correlation Strength

Mainstream ISO tar, nicotine, and carbon monoxide were strongly correlated with each other and with most mainstream constituents, as indicated by R^2. There were exceptions, however. ISO fundamental yield parameters were weakly correlated with the yields of nitric oxide, nitrogen oxides, and TSNA with $R^2 < 0.80$. As will be seen later in this text, tobacco chemistry can be an important variable to consider with these smoke constituents for improving predicting models. Smoke pH was relatively constant among brands. This resulted in very weak correlations and a nearly horizontal regression line. This same trend of near-constant smoke pH for various brand styles was seen in other studies (Massachusetts, 1999, 2000, 2001).

Mainstream smoke constituents tend to be associated more with either the particulate phase or the vapor phase. The distinction is not absolute (Hoffmann, 1993). Nicotine, polycyclic aromatic hydrocarbons, and TSNA are examples of constituents associated with the particulate phase and, as such, retained by a standard Cambridge filter pad (Wartman et al, 1959). Lower molecular weight hydrocarbons such as benzene, 1,3-butadiene, and toluene are primarily vapor phase constituents. Hydrogen cyanide and ammonia for example, are found both in the vapor and particulate phases. The significant amount of mainstream smoke formaldehyde retained by the Cambridge filter has been attributed to its hydrophilic nature and subsequent association with total particulate matter on the

pad (Spencer and Chard, 1971). From the statistics in Table 5, ISO tar is a reasonable parameter choice for linear correlations of ISO constituent yields. This choice is based on regression statistics for constituents associated with either the vapor or particulate phase. In general, ISO nicotine and carbon monoxide demonstrated marginally weaker association, and therefore poorer predictive ability, with vapor and particulate phase constituents. For this study, ISO tar was selected as the regressor for functional relationship development and subsequent prediction. It was noted before that tar, nicotine and carbon monoxide had been used to describe the functional relationships for smoke constituents in a benchmark study for the Massachusetts DPH (Borgerding et al., 2000). Others involved with cigarette development, consumer exposure research, or regulatory concerns may have different analysis needs. Carbon monoxide or nicotine would also be appropriate regressors for describing the yield relationships for smoke constituents as both nicotine and carbon monoxide yields were highly correlated with tar yields.

(3) Prediction Error

To test the reliability of the established linear functional relationships, the measured constituent yields for the nine validation brands were compared with the predicted yields and the 95% prediction intervals. Table 6 shows the prediction errors (PE) using ISO tar. The majority (>90%) of constituents in the validation brands were within the 95% prediction intervals. Measured brand average constituent yields outside the prediction intervals are highlighted in Table 6. No constituents consistently fell outside of the prediction intervals

across the validation brands. Two brands containing only bright tobacco (V2 and V7) tended to be under-predicted for several constituents. Differences in cigarette smoke composition between all bright and all burley cigarettes have been reported (Griest and Guerin, 1977). It was observed that mainstream smoke yields for phenol, cresols, isoprene, and benzo[a]pyrene were higher for bright cigarettes than for burley cigarettes. Mainstream smoke from all bright cigarettes was reported to be lower in nitrogen oxides and higher in hydrogen cyanide. In addition to tobacco blend differences, there are differences in processed tobacco components in different market regions. Tobaccos from different regions will be impacted further by agricultural practices and growing conditions. For example, the concentration of metals in tobacco, and consequently smoke, is affected by agricultural practices and growing conditions (Stohs et al., 1997, and Smith 1997). Smoke cadmium yield for brand style (V8) in the study was beyond the prediction interval. This could be the result of unusual growing conditions for a particular crop year or years. Although tobacco blending over several crop years tends to mitigate year-to-year differences, any uncharacteristic change in yield would not be predicted in a benchmark study. One underlying premise of benchmarking is that brands are fairly consistent over time. Swauger et al. (2002) compared smoke constituent yields for over 30 U.S. commercial cigarette brands for sampling years 1995, 1998, and 2000. Among their findings was smoke constituent yields of commercial U.S. cigarettes had remained fairly constant between 1995 and 2000.

The PE's for each constituent in Table 6 are comparable in magnitude to the standard deviations of measured yields of the validation brands in most of the cases. Exploratory brand average yields, best-fit lines, 95% prediction intervals, and validation group average yields for each smoke constituent are shown in Figure 1. To exemplify the inherent variability, one standard deviation error bars are included for several constituents. Standard deviation data for all of the constituents are included in Table 3.

(4) Brands with Carbon in Filter Design

Nine brands in the exploratory group contained carbon in their filter construction. It is well known that carbon, i.e., activated charcoal, has the potential to selectively adsorb many vapor phase constituents from mainstream smoke (Baggett and Morie, 1975, and Williamson et al., 1965). Carbon monoxide and oxides of nitrogen are generally not affected by carbon (Tiggelbeck, 1972). Brands with carbon-containing filters in this study therefore would be expected to influence the functional regression relationships of vapor phase constituents with a bias toward lower predicted yields for non-carbon filter brands. Similarly, some vapor phase constituents would be over-predicted for brands with filters containing carbon. This influence was reflected in both the quality of the regression relationships (regression diagnostic statistics) and prediction interval width. Figure 2 demonstrates the differences between brands with carbon in filters and those without for benzene yields. The lack of carbon influence on particulate phase benzo[a]pyrene yield, for example, is also seen in this figure.

Functional relationship models for vapor phase constituents were consequently expanded to include a factor for the presence of carbon in filters. Multiple regression relationships were developed using tar and an "indicator" variable of 1 or 0 for the presence or absence of carbon, respectively, for the 39 exploratory brands. Again, regressions were weighted with the inverse of tar yield. The results showed general reductions in root mean square errors and increases in R^2 for most of the vapor phase constituents. Diagnostic statistics for weighted regression, with and without the "indicator" variable, are compared in Table 7. Use of an "indicator" variable for filters with carbon does not take into account different filter construction, different weights of carbon in carbon filters, or the surface areas and activities of carbons used in filters for various brands. Multiple regression adds a level of complexity to both relationship development and application of those relationships for predicting yields of other brands. The potential impact of carbon filters should be considered in assessing brand constituent yields and predicting relationships for vapor phase constituents. The benchmarking approach would be equally valid if brands with carbon in filters were modeled separately from those without carbon in future studies. Figure 3 is a comparison between measured smoke yields and linear and multiple regression predicted yields for benzene and acrolein, as examples. With simple linear regression, there is one predicted yield line. Predicted yields from multiple regression with tar and a carbon filter factor are not on a single line. The lower predicted yields are for brands with carbon, as the coefficient for the carbon indicator variable was always negative. It is seen that there is less over-

prediction (predicted yield higher than observed) for yields below the linear model and less under-prediction (predicted yields less than observed) of yields above the linear model line with multiple regression. Predicted yields for brands without carbon were greater than predicted yields using linear regression with carbon only. In general, multiple regression with a carbon factor resulted in lower prediction errors for vapor phase constituents of most brands.

(5) Mainstream TSNA and Nitrogen Oxides Treatment

Many of the chemical constituents found in smoke are generated during combustion from precursors found in the leaf. For example, phenolic compounds are derived partially from cellulosic materials in leaf, some nitrogen compounds are derived partially from proteins in leaf, nitrogen oxides are derived partially from nitrates in leaf, and carbon oxides come from many leaf precursors through combustion (Baker, 1999). The amounts of these constituents found in mainstream smoke depend in part on the amounts of their precursors in the tobacco that is consumed during puffing and the conversion efficiencies of the precursors. Other smoke constituents are distilled intact from the tobacco to smoke during the smoking process. Mainstream smoke constituents such as nicotine (Gorrod et al, 1999) and, to some extent, TSNA are in this category (Fischer et al, 1990). The amounts of these constituents found in mainstream smoke depend on their concentrations in the tobacco that is consumed during puffing and their transfer efficiencies to smoke.

As noted earlier, nitrogen oxides and TSNA were weakly correlated with ISO tar, nicotine, or carbon monoxide. Tobacco filler nitrate and TSNA

concentrations were used to modify functional relationships for predicting the yields of nitrogen oxides and TSNA in mainstream smoke. Tobacco filler concentrations of nitrate and TSNA for each brand style in this study are found in Table 8.

For TSNA, multiple regression with ISO tar and filler TSNA concentration was first assessed. The TSNA cigarette concentration was calculated from the dry-weight basis (less water content) cigarette tobacco weight and the dry-weight basis concentration of the component in tobacco filler. Multiple regression relationships for NNN, NNK, NAT, and NAB were significantly improved over simple linear regression models as evidenced by higher respective R^2's, lower root mean square errors, and lower prediction errors for both the exploratory and validation brands. Examination of multiple regression residuals revealed that brands lower in both tar yield and filler TSNA concentration tended to be under-predicted. Brands with either higher tar yields and lower tobacco filler TSNA concentration or brands with lower tar yields and higher tobacco filler TSNA concentration tended to be over-predicted. For this reason, variable transformations were investigated. Square root (sqrt) transformed primary and interactive variables, sqrt(tar), sqrt(cigarette filler TSNA), and sqrt(tar x cigarette filler TNSA), were regressed with smoke TSNA yield. The coefficient for the second primary variable, sqrt(cigarette filler TSNA), was statistically insignificant. The coefficient for the first primary variable, sqrt(tar), was considerably less significant than the coefficient for the interactive variable. The final relationship used for prediction was a simple linear regression of smoke TSNA yields with the

E. Kaelin 5/21/03

not significant.

transformed, interactive variable sqrt(tar x cigarette filler TSNA). Residuals were homogeneously and normally distributed over the transformed variable range. The result was a significant increase in regression quality with higher R^2 and lower prediction errors for all of the brands in the validation set. This simple model also allowed for one-dimensional prediction intervals and simpler visualization. NAB was an exception, however. Differences between multiple and simple regressions with transformed variables were minimal. Of the four TSNA analyzed, prediction errors were relatively greater for NAB. Figure 4 contains transformed variable regression models and prediction intervals for mainstream smoke NNN, NNK, NAT, and NAB. A similar approach was used for regression model development for nitric oxide and nitrogen oxide yields. Prediction model quality between multiple regression using tar and cigarette nitrate concentration and simple linear regression using a square root transformed interactive term variable, sqrt(tar yield x cigarette nitrate concentration) was similar. Prediction models for nitric oxide and nitrogen oxides are found in Figure 5. Prediction errors were reduced relative to simple linear regression using only tar as the independent variable. The use of prediction models beyond simple linear regression for TSNA or nitrogen oxides would have less impact if all brands were more similar in tobacco filler nitrate or TSNA concentration. Since this study included brands from a wide range of regional markets, and consequently a wide range of concentrations of these tobacco components, different regression models were warranted.

(6) Relative Prediction Errors

To capture the overall impact of the different regression models on prediction errors for the validation brands, an average absolute relative prediction error (average ARPE) was calculated for each constituent. This was simply the sum of absolute relative prediction errors for each brand averaged over the nine validation brands. The relative prediction error for a constituent yield in a particular validation brand is the difference in measured and predicted yields divided by the measured yield. The average ARPE was then compared to the coefficient of variation, %CV, for each constituent averaged over the nine validation brands. These averaged values are found in Table 9. For vapor and particulate phase constituents, prediction errors were generally within two standard deviations of the measured yield. Average ARPE's were therefore approximately two times the respective average %CV. It should be recognized, with smoke constituents measured at the microgram and nanogram levels, that very small differences in predicted and measured yields will result in large ARPE's. For example, the average ARPE for the nine brands for phenol and pyridine were 22% and 21%, respectively. Those percentages would equate to average prediction errors of 2.1 micrograms and 1.8 micrograms. An ARPE of 21% for NNK yields would equate to an average prediction error of 10 nanograms. If the average predicted error for 4-aminobiphenyl were 1.0 nanograms, the ARPE would be 50%.

DISCUSSION

A benchmarking approach for predicting reliable cigarette smoke constituent yield information was evaluated with a worldwide sample of commercial brands. The approach was applied to smoke yield data obtained at standard ISO smoking conditions. Results for the diverse range of brands supported the underlying assumption that smoke constituent precursors in tobacco were available somewhat uniformly across tobacco blends from the different regions. Precursors include, for example, tobacco proteins, cellulose and lignin materials, and nitrogen-containing compounds. Mainstream smoke constituent yields increased with smoke tar, nicotine, or carbon monoxide yields. Smoke tar is directly related to tobacco consumed during puffing, which in turn is dependent on cigarette design.

Benchmarking is based on two key elements. Cigarettes for which smoke constituent yield predictions are to be made must have design and tobacco blend characteristics within the boundary established by the benchmarking set of test brands. The boundary is defined by the range of the cigarette design characteristics of the test brands. This particular study was for Philip Morris brands manufactured in various regions. Extrapolation to other brands for which brand design, quality, and consistency are not known could be misuse of the regression relationships. Interpolation beyond the design range could lead to predicted yields with errors of unknown magnitude. The second key element for benchmarking is the single-laboratory study design. Since there are currently no universally recognized smoke constituent analyses methods, it was important to

use the same laboratory and methods for a particular constituent analysis. The single laboratory approach in which quality controlled and validated methods are used is a viable option for maintaining data consistency. However, the accuracy of the test methods in a single laboratory approach can not be determined in the absence of standardized methods.

The benchmark approach utility is determined by how closely predicted yields compare to measured yields. Prediction errors for smoke constituents of the validation brands in this study were moderate when examined relative to other sources of potential variation. One source is the point-in-time variation in a single set of sample replicates, as evidenced by the magnitude of the analysis standard deviation. Additional variation is inherent in more subtle day-to-day fluctuations in instrument calibrations, technician performance, and smoking machine operations. Testing laboratories practicing sound quality control procedures will minimize, but not eliminate, these analysis variations. Variability in method execution over time was estimated by comparing average constituent yields for the 1R4F Kentucky reference cigarette reported by the contract lab for several company studies (Table 10). Analyses were performed over an 11-month period with the laboratory using the same methods throughout. Consistency of the reference cigarette was assumed. Average 95% confidence intervals for standard smoke parameters (tar, nicotine, puff count, etc.) were approximately +/- 7% of parameter means. For mainstream smoke constituents, the average 95% confidence intervals for the mean yields were considerably wider. Vapor phase and particulate phase constituent average 95% confidence

intervals were +/- 18% and 22% of mean yields, respectively. These confidence intervals were of similar magnitude to most of the prediction errors in the current benchmark study.

A second source of variation in smoke constituent yields is the cigarette brand itself. Three potential sources of cigarette variability are described in Annex C of ISO Standard 8243 (ISO, 1991). Short-term variability is attributed to minor fluctuations in tobacco blending and weight, and non-tobacco components (e.g., paper porosity and filter ventilation) during manufacture. Medium-term variability reflects short-term variability and also batch-to-batch changes in non-tobacco components, tobacco grades, and machinery performance. Long-term variability can result from crop year differences or machinery replacement, for example, and are additive with short- and medium-term variability. As previously noted, natural products such as tobacco are subject to crop year weather variations, regional agricultural practices, and tobacco processing differences. Tobacco blending from different crop years minimizes these differences, but some medium- and long-term variation is expected. For this reason, a smoke constituent yield measured for a brand produced today could vary from the yield for the same brand produced at another time. An example of typical smoke constituent yields variation for a single brand produced over a period of time is seen in Table 11. The brand was tested monthly over a nine-month period by an internal research laboratory. The measured yield variation encompasses short- and medium-term brand variation and laboratory method variation over time. A significant point of the information in Table 11 is the magnitude of the 95%

confidence intervals around the mean constituent yields. The average 95% confidence interval for all smoke constituents analyzed was +/- 26% of the means for this time period. These intervals are only moderately greater than the confidence intervals around the means for the 1R4F reference cigarette tested by the contract laboratory over time. The intervals are also similar in magnitude to the average relative prediction errors for the validation brands in this current benchmark study. These two examples of smoke constituent yield variations support the premise that the predicted yields and prediction intervals, generated from an appropriately designed benchmark study, are valid options for providing smoke constituent yield information for conventional cigarette brands represented in that study design. The 95% prediction intervals generated from the regression functions provide reasonable expectations of the variation for brands at some average tar, nicotine, or carbon monoxide yield level.

Benchmarking utility will be compromised if particular constituents were analyzed in different laboratories without method standardization. Purkis, et al. (2001) compared ISO smoke constituent yields for three brands tested by six laboratories. Each laboratory used its own internally validated methods. All laboratories reported the same relative ranking of constituents yields for the brands but there was significant variation between laboratories for particular constituent yields. Ratios of highest to lowest reported yields ranged from 1.09 (tar) to 3.20 (quinoline). The authors postulated that, on average, differences greater than 60% between laboratories would be required for confidence that observed yield differences were statistically significant. In contrast, within-

laboratory variation for constituent yields was significantly lower with average coefficients of variation of 9%. Data reported in Table 12 reinforce the conclusions of Purkis. Smoke constituent yields for the 1R4F reference cigarette reported by two contract laboratories (point-in-time measurement) and from several literature sources were compared. The ratios of maximum to minimum reported constituent yields ranged from moderate (approximately 1.1) to significant (greater than 2.0). These examples illustrate the need for caution in comparing smoke constituent yield data from different testing sources.

A benchmark approach can provide predicted smoke constituent yield information as demonstrated by this study. A limitation of benchmarking is the inability to reliably predict constituent yields for brands with some unusual characteristic not anticipated from the design range of the model brands. This limitation would be minimized by successively adding new brands to the benchmarking model brands over time and including a validation set of brands as an internal check of the functional relationships. Benchmarking is useful only to the extent that the test brands are representative of brands for which yield predictions are to be made. Prediction errors cannot be eliminated. The magnitudes, however, could be minimized provided the basic elements of a benchmarking approach are recognized. These basic elements include: (1) test cigarette brands have design features representative of the marketplace; (2) the predicted brands have design features within the ranges defined by the test brands; (3) brand quality is consistent over time; and (4) each constituent is

analyzed by a single laboratory until method standardization across laboratories is achieved.